UDC 551.577.36:551.579.5:625.72:625.042:65.015.2

Forecasting Work Conditions for Road Construction Activities: An Application of Alternative Probability Models

EMIL D. ATTANASI and S. R. JOHNSON—*Department of Economics, University of Missouri, Columbia, Mo.*

SHARON LeDUC—*Department of Atmospheric Science, University of Missouri, Columbia, Mo.*

JAMES D. McQUIGG—*Environmental Data Service, NOAA, and Department of Atmospheric Science, University of Missouri, Columbia, Mo.*

ABSTRACT—Certain activities of highway construction are particularly sensitive to such weather conditions as soil moisture, precipitation, and daily temperature. Regression analysis is used to obtain three alternative probability models designed to translate observed weather conditions into probabilities for carrying out construction activities. The models were developed using generalized least squares, normit analysis, and logit analysis. The generalized least squares method was the most convenient computationally, but it had severe interpretative disadvantages. The results obtained by logit analysis gave the desired probabilistic interpretation most readily and had the best predictive ability. Comparison of sample observation and predicted work probabilities for common excavation during wet and dry months indicated that the logit analysis model could accurately translate weather conditions into probabilities that work would take place. Models for paving and asphalt work and for bridge and drainage structure are also estimated using logit analysis. These estimates indicate a strong sensitivity of the latter category of work to precipitation conditions. Such models may aid contract letting agencies in planning payment schedules, penalty clauses, and completion dates for new roads; construction firms may find such models valuable in planning effective use of men and equipment.

1. INTRODUCTION

The road construction industry is highly sensitive to weather conditions. This sensitivity arises from the fact that certain types of weather conditions must exist for the various construction activities to operate efficiently. For example, common excavation cannot be efficiently carried out during periods of excessive rainfall, concrete pouring is hazardous during periods in which the temperatures are below freezing, and the like. Because of this sensitivity to ambient weather conditions, it is important for contractors and contracting firms to have a convenient means by which to translate weather forecasts and observations into probabilistic statements regarding the feasibility of conducting the various road building activities.[1]

In regard to forecasts, these probabilistic statements can aid contracting firms in deciding whether or not to ask laborers to report to work. Such forecasts are of increased importance if union arrangements attach a high penalty to errors in such decisions as they currently do; that is, if union contracts require partial or full wages for days in which laborers are called out but work is not feasible. Translation of observed weather conditions into probabilities for conducting construction activities is also important in long-range planning. With such information, one can use historical weather records to calculate the number of days in which various construction activities would have been possible in previous years. The historical perspective gained from these results could help contract-letting agencies plan payment schedules, penalty clauses, and opening dates for new highways; at the same time, construction firms could more effectively plan the use of men and equipment and, thereby more accurately determine their capacity for taking on new contracts.[2]

One method for analytically determining work conditions from sample data is based on the idea of a linear probability model. Regression techniques are used to relate variables reflecting actual working conditions to available information on climatic variables at construction sites. In this paper, we apply the probability models to data made available by the Missouri State Highway Commission. The objective is to show how these probability models can translate weather data into useful information for determining the daily feasibility of alternative construction activities and to indicate how such information might be used.

[1] State Highway Departments typically use district engineers to make onsite decisions as to whether or not conditions are suitable for work. Since penalty clauses are based both on numbers of working days and numbers of calendar days, variability from one district engineer to another frequently causes difficulties in decisions concerning work conditions.

[2] Contractors and contracting agencies currently have methods of translating their experience into operational and planning decisions. The proposed methods and results are designed to complement existing procedures by facilitating the use of existing weather data in obtaining additional information to be used in such decision-making processes.

2. SPECIFICATION AND ESTIMATION OF PROBABILITY MODELS

First, it is necessary to introduce a mechanism for quantifying working conditions. Initially, two classifications of work conditions will be used: (1) work day, where all conditions are suitable for a particular construction activity, and (2) no-work day, where some or all conditions are unsuitable for a particular construction activity. Given these two classifications, we can introduce an artificial variable for quantification. The artificial variable to be employed has the value 1 for work days and 0 for no-work days.

As previously indicated, the probability model is designed to estimate probabilities of work days on the basis of observable climatic variables. These probabilities are estimated from a random sample of size N of concomitant observations of the artificial variable and the climatic variables. More formally, let y_n be the nth value of the observed artificial variable, X'_n be the nth observation on the set (e.g., k) of climatic variables, and ξ_n be the nth disturbance term. With this specification of the variables, the objective of the probability models is to determine a parameter vector $\boldsymbol{\beta}$, conformable with X'_n, which will give estimated values of y_n with some statistically desirable properties. In other words, we wish to estimate $\boldsymbol{\beta}$ from the model,

$$\mathbf{y}=\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\xi}. \tag{1}$$

In specifying eq (1), we have used $\mathbf{y}$ to denote the vector of N sample values for y_n, $\mathbf{X}$ to denote the N sample values with X'_n as rows, $\boldsymbol{\xi}$ as the associated vector of unobservable disturbance terms, and the prime mark to denote the transposition of a matrix.

Problems encountered in obtaining estimates of $\mathbf{y}$ from eq (1) are mainly associated with peculiarities of the disturbance term, $\boldsymbol{\xi}$. These peculiarities, as we shall subsequently demonstrate, come about as a result of $\mathbf{y}$ being a qualitatively specified variable.

In the discussion to follow, we develop three procedures for estimating $\boldsymbol{\beta}$ from eq (1). These estimation procedures are heuristically developed in connection with ordinary least squares. The approach serves to tie the procedures to the more customary methods of estimation; this is accomplished at little expense in terms of space. It also points up the fact that the proposed estimation procedures are derived on the basis of the treatment of the disturbance term, $\boldsymbol{\xi}$.

Given a model of the form of eq (1), one naturally contemplates the estimation of $\boldsymbol{\beta}$ by ordinary least squares. This approach has some decided limitations, however. Application of ordinary least squares requires $E(\mathbf{X}'\boldsymbol{\xi})=0$, $E(\boldsymbol{\xi})=0$, and $E(\boldsymbol{\xi}\boldsymbol{\xi}')=\boldsymbol{\Omega}=\sigma^2\mathbf{I}$, where $\mathbf{I}$ is the identity matrix and σ is a scalar constant, if estimated parameter values are to be unbiased and efficient.[3] The inapplicability of these conditions is easily shown. If we again denote the nth sample value of $\mathbf{X}'$ as X'_n, it follows that $\xi_n=y_n-X'_n\boldsymbol{\beta}$. Since y_n is either 0 or 1, ξ_n must be either $-X'_n\boldsymbol{\beta}$ or $1-X'_n\boldsymbol{\beta}$. Hence, for $E(\boldsymbol{\xi})=0$, $E(\xi_n)$ must be $1-X'_n\boldsymbol{\beta}$ when $\xi_n=-X'_n\boldsymbol{\beta}$ and $X'_n\boldsymbol{\beta}$ when $\xi_n=1-X'_n\boldsymbol{\beta}$. On these conditions, the variance of ξ_n, $E(\xi_n^2)$, is equal to $E(y_n)[1-E(y_n)]$. Thus, the disturbance varies with $E(y_n)$ or equivalently with the observed climatic variables, X'_n, and $E(\boldsymbol{\xi}\boldsymbol{\xi}')\neq\sigma^2\mathbf{I}$.[4]

The appropriate estimation procedure for situations in which $\boldsymbol{\Omega}\neq\sigma^2\mathbf{I}$ is the Aitken's generalized least squares (Goldberger 1964, p. 233), rather than the ordinary least-squares procedure. Although the generalized least-squares estimators have some undesirable properties in terms of estimated values of y_n, they are more efficient than the ordinary least-squares estimators and are easy to compute. On the basis of these latter two properties, the generalized least-squares estimators are suggested as one of the methods appropriate for estimating work day probabilities. Using the generalized least-squares method, we can estimate $\boldsymbol{\beta}$ from eq (1) by (1) estimating $\boldsymbol{\Omega}$ from the ordinary least-squares residuals and (2) using the estimated value, $\hat{\boldsymbol{\Omega}}$, in obtaining an estimation of $\boldsymbol{\beta}$, as

$$\hat{\mathbf{b}}=(\mathbf{X}'\hat{\boldsymbol{\Omega}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\boldsymbol{\Omega}}^{-1}\mathbf{y}. \tag{2}$$

The estimated variance-covariance matrix for the generalized least-squares estimator, $\hat{\mathbf{b}}$, is $(\mathbf{X}'\hat{\boldsymbol{\Omega}}^{-1}\mathbf{X})^{-1}$. Then $\hat{y}_n=X'_n\hat{\mathbf{b}}$ is naturally interpreted as the conditional probability of a work day, given that the climatic conditions described by X'_n occur.

It is in regard to the estimated values for y_n that the limitations of the Aitken's estimators are found. These limitations exist because the procedure incorporates no restriction that the $\hat{y}_n$ fall within the unit interval. Therefore, although the generalized least-squares estimators may yield good first approximations of the conditional probabilities, one may still obtain values for $\hat{y}_n$ that are outside the 0, 1 interval.

Two similar estimation methods are designed to handle this difficulty. As might be expected, these methods involve the imposition of further restrictions or conditions on ξ_n. The proposed methods are called normit and probit analyses (Berkson 1955, Goldberger 1964, and Tobin 1955). Since the two procedures involve similar methods for handling the problem of $\mathbf{y}$ outside the 0, 1 interval, their derivations are highly related. We will focus upon the normit model, keeping in mind that it can be slightly altered to obtain the probit model.

The normit estimation procedure is based on the following considerations. Let I_n be defined as a linear function of the regressors; that is, $I_n=X'_n\boldsymbol{\beta}$. Also, let I^*_n be distributed $N(0, 1)$ and assume that the value of y_n is given by

$$y_n=\begin{cases}1 \text{ if } I_n\geq I^*_n\\ 0 \text{ if } I_n<I^*_n.\end{cases} \tag{3}$$

Notice that this assumption simply involves a set of conditions on the ξ_n in eq (1). It follows from eq (3) that the y_n are now functions of the X'_n and I^*_n. The I^*_ns, which are

[3] E is the expectation operation. The unbaised and efficiency conditions follow from the Gauss-Markov theorem on least squares.

[4] Goldberger (1964, pp. 248–252) has a more elaborate discussion of these details.

in essence disturbances, are interpreted as critical values of the index I_n. It follows from eq (3) that, if we let $F(z)$ denote the value of the cumulative of the standard normal distribution at z, then

$$P = \text{Prob.}\ (y_n=1|I_n) = \text{Prob.}\ (I_n^* \leq I_n|I_n) = F(I_n) \quad (4)$$

and

$$Q = \text{Prob.}\ (y_n=0|I_n) = \text{Prob.}\ (I_n^* > I_n|I_n) = 1 - F(I_n) \quad (5)$$

where the equivalence of these conditional statements follows from the assumed $N(0, 1)$, distribution of I_n (Zellner and Lee 1965, p. 384).[5] Since I_n (and thus the probabilities P and Q) are functions of $\boldsymbol{\beta}$ (through the definition of I_n^*), a maximum likelihood or minimum chi-square procedure can be employed in estimating the parameters of the linear functions defining I_n^*.

The asymptotic properties of the maximum likelihood and minimum chi-square estimations are the same. The minimum chi-square procedure is developed here because it is computationally more convenient. Following the development of Zellner and Lee (1965, pp. 383–385), we can combine eq (3) and (1) (recalling, of course, that we are simply redefining ξ) to get

$$F^{-1}(P_n+\xi_n) = F^{-1}(P_n) + \xi_n \frac{dF^{-1}}{d(P_n)} + R_n \quad (6)$$

where F^{-1} is the inverse of the cumulative normal distribution; the right side of eq (6) is an expansion about $X_n'\boldsymbol{\beta}$ with R_n representing higher order terms. Differentiating eq (6) with respect to the distribution function and using the definition of I_n, we have

$$I_n^o = X_n'\boldsymbol{\beta} + \frac{\xi_n}{Z(P_n)} + R_n. \quad (7)$$

I_n^o is the observed "normit" and $X_n'\boldsymbol{\beta}$ is the "true" normit while $Z(P_n)$ is the value of the unit normal.

The variance of the newly defined error term $\xi_n/Z(P_n)$ is given by

$$\text{Var}\left[\frac{\xi_n}{Z(P_n)}\right] = \frac{P_n(1P_n)}{[Z(P_n)]^2}. \quad (8)$$

Hence, $\boldsymbol{\beta}$ can be estimated by minimizing the normit χ^2, where

$$\text{Normit}\ \chi^2 = \sum_{n=1}^{N} \frac{[Z(P_n)]^2}{(P_n+\xi_n)(1-P_n+\xi_n)} (I_n^o - X_n'\boldsymbol{\beta})^2, \quad (9)$$

with respect to the vector $\boldsymbol{\beta}$ [eq (1)]. The procedure for obtaining $\boldsymbol{\beta}$ reduces to an estimator similar to the Aitken's estimator but with the "weights" given by $\boldsymbol{\Omega}_e$, a diagonal matrix composed of the terms post-multiplied by $(y_n - X_n'\boldsymbol{\beta})$ in eq (9). For purposes of the subsequent empirical work, the elements of $\boldsymbol{\Omega}_e$ are approximated by using the least-squares estimators of the vector $\boldsymbol{\beta}$ in computing the weights. Specifically, the term $Z(X_n'\mathbf{b})/X_n'\mathbf{b}(1-X_n'\mathbf{b})$ is employed in estimating the weights, with $\mathbf{b}=(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ and the additional conditions that the arguments in the weights are set to 0.99 for n when $X_n'\mathbf{b} \geq 1$ and 0.01 for n when $X_n'\mathbf{b} \leq 0$. The estimated probability of a working day, given that the climatic conditions X_n' prevail, is then

$$P = F(\tilde{I}_n) = F(X_n'\tilde{\mathbf{b}}) = F[X_n'(\mathbf{X}'\boldsymbol{\Omega}_e^{-1}\mathbf{X})(\mathbf{X}'\boldsymbol{\Omega}_e^{-1}\mathbf{y})] \quad (10)$$

where $\tilde{I}_n$ is the sample estimate of the true normit and $\tilde{\mathbf{b}}$ is the sample estimate of $\boldsymbol{\beta}$.

A second approach to the problem of obtaining estimated values for probabilities that lie outside the unit interval is called the logit model. The estimation procedure associated with the logit model is, in principle, similar to that of the normit model. To demonstrate the derivation of the procedure, we again write an n index based upon fixed variables X_n'; for example, $J = X_n'\boldsymbol{\beta}$, where $n=1, 2, \ldots, N$. We suppose that the probability that $y_n=1$ can be written

$$\text{Prob}\ (y_n=1|X_n') = F(X_n'\boldsymbol{\beta}) = \frac{e^{X_n'\beta}}{1+e^{X_n'\beta}}. \quad (11)$$

This expression can be derived on the basis of the usual discriminant analysis formulation of the problem. In fact, the expression is the ratio of probability distributions used in the initial phase of discriminant analysis. However, since a model based on the set of fixed explanatory variables X_n', where $n=1, \ldots, N$, is now being postulated, the function will not be simply fitted to the discriminant function (as in standard discriminant analysis). Instead, we estimate the parameters or weights using a different approach and, thus, obtain estimators with different properties. As with the normit analysis, we can form a likelihood function for the parameters to be estimated, given eq (11) and a similar expression for the probability that $y_n=0$.

In deriving the likelihood function, we assumed that the y_n are ordered so that the first s are ones and the second $N-s$ are zeros. The likelihood function for the sample can then be written

$$L^* = \left(\Pi_{n=1}^{s} \frac{e^{X_n'\beta}}{1+e^{X_n'\beta}}\right)\left(\Pi_{n=s+1}^{N} \frac{1}{1+e^{X_n'\beta}}\right). \quad (12)$$

Taking logarithms and differentiating with respect to $\boldsymbol{\beta}$, we obtain

$$\frac{\partial \ln L^*}{\partial \beta_i} = \sum_{n=1}^{s} X_{ni} - \sum_{n=1}^{N} \frac{X_{ni}e^{X_n'\beta}}{1+e^{X_n'\beta}} = 0, \quad (13)$$

or, utilizing eq (11), we get

$$\sum_{n=1}^{s} X_{ni} - \sum_{n=1}^{N} X_{ni}\ \text{Prob}\ (y_n=1|X_n') = 0. \quad (14)$$

The logit procedure produces a set of normal equations that must be solved for the parameters by iterative processes.

[5] Specifically, it follows from our assumptions and the definitions implicit in eq (4)-(6) that

$$F(I) = (2\pi)^{-1/2} \int_{-\infty}^{X_n'\beta} \exp\left(-\frac{u^2}{2}\right) du.$$

The computational procedure may be greatly simplified by using the approximation

$$\frac{e^j}{1+e^j}=\begin{cases}1 & j>3\\ \frac{1}{2}+\frac{1}{6}\cdot & j-3\leq j\leq 3\\ 0 & j<-3\end{cases} \tag{15}$$

The procedure used in this paper and developed is due to Cox (1966, pp. 60–61). To obtain initial estimates of the βs, we used eq (15) and applied unweighted least squares to the model,

$$E(y_i)=\frac{1}{2}+\frac{1}{6}(\hat{\beta}_0X_0+\cdots+\hat{\beta}_kX_k)\cdot \tag{16}$$

Recalling that we have k climatic variables, we define $\boldsymbol{\beta}$ to be a $(k+1)\times 1$ vector of estimates, $\mathbf{S}_0$ to be a $(k+1)\times(k+1)$ matrix of cross-products for which the (r,s) element is $\sum_i X_{i_r}X_{i_r}$, and $\mathbf{T}_0$ to be a $(k+1)\times 1$ column vector for which the rth element is $\sum_i(y_i-\frac{1}{2})X_{i_r}$. This first approximation $\tilde{\boldsymbol{\beta}}$ satisfies

$$\frac{1}{6}\mathbf{S}_0\tilde{\boldsymbol{\beta}}=\mathbf{T}_0\cdot \tag{17}$$

We adjust for the discontinuity in the approximation obtained by eq (17) by the following procedure. For observations such that

$$|\tilde{\beta}_0X_{i0}+\ldots+\tilde{\beta}_kX_{ik}|>3, \tag{18}$$

we delete their contribution to $\mathbf{S}_0$ giving a new matrix $\mathbf{S}_1$. Also, $\mathbf{T}_0$ is redefined to be $\mathbf{T}_1$ for which the rth element is

$$\sum_i\left(y_i-\frac{1}{2}\right)X_{ir}+\frac{1}{2}\left(\sum_i{}^{-}X_{ir}-\sum_i{}^{+}X_{ir}\right) \tag{19}$$

where Σ^- and Σ^+ denote summations over these observations such that,

$$\text{and}\quad\begin{aligned}\textstyle\sum^-&=\tilde{\beta}_0X_{i0}+\ldots+\tilde{\beta}_kX_{ik}<-3\\ \textstyle\sum^+&=\tilde{\beta}_0X_{i0}+\ldots+\tilde{\beta}_kX_{ik}>3.\end{aligned} \tag{20}$$

The newly constructed $\mathbf{S}_1$ and $\mathbf{T}_1$ matrices can be used to obtain maximum likelihood estimates of $\boldsymbol{\beta}$ for the second iteration ($\approx$) from

$$\frac{1}{6}\mathbf{S}_1\tilde{\tilde{\boldsymbol{\beta}}}=\mathbf{T}_1. \tag{21}$$

This procedure may be applied iteratively to refine estimates of the $\boldsymbol{\beta}$s to satisfy eq (14).

Note that the logit and normit models are similar. In both, a "symmetrical sigmoid curve" is fitted to a linear function of the observed data sample. The computational burden of the logit model is not as severe as that for the normit analysis because of the degree of nonlinearity in the first-order conditions. In particular, the normit analysis requires the evalution of a number of normal integrals. Since both methods require an assumption about the distribution form for the two populations, neither has much advantage in terms of statistical generality. Although both types of estimations for the weights on the weather variables will be subsequently presented, applications of these approaches for operational purposes should probably be guided by computational feasibility. On the basis of this computational consideration, approximations based on the generalized least squares may be the most feasible. The subsequent comparisons of results for the Missouri Highway Department data may provide some insights into the legitimacy of this last observation.

3. DATA FOR CONSTRUCTION AND CLIMATIC VARIABLES

Records giving the sample values for the qualitative variables, y_n, were obtained from the Missouri State Highway Commission. These records were taken from two highway construction projects completed in the vicinity of Jefferson City, Mo., during 1966–67.[6] Specifically, the records included information on common excavation, finishing and grading, and paving activities. Although more detailed information on each of these activities was available, our data were supplied by the engineers' reports indicating when the activities were active and when they were inactive. For purposes of maintaining some homogeneity in the data, the sample days were restricted to those occurring between April 1 and October 31. The records from the two projects gave 218 observations for common excavation, 122 observations for bridge, culvert, and drainage structures, and 80 observations for paving and asphalt. Because of the relative number of observations on common excavation, the importance of common excavation in road building, and the sensitivity of common excavation to climatic conditions, the major portion of the numerical results will be related to this particular construction activity.

As was implied earlier, the independent variables are related to climatic conditions. They are X_{1n}, a soil mositure index; X_{2n}, a 7-day average precipitation; X_{3n}, 4-day average precipitation; X_{4n}, precipitation for the current day plus X_{7n}; X_{5n}, a 0, 1 variable indicating whether or not precipitation occurred on the day on which working conditions were recorded; X_{6n}, the average temperature on the day the work conditions were recorded; and X_{7n}, the 3-day average precipitation. In the variables defined as averages, the period includes the calendar days immediately prior to the working day in question.

With the exception of X_{1n}, the climatic variables are all available in the records for cooperative weather stations. In the case of this study, they were taken from the National Weather Service Climatological Station located at Lincoln University, Jefferson City, Mo., approximately 2 mi from the construction sites. The soil moisture variable is derived from precipitation and temperature data. It is designed to reflect the moisture content in the top

[6] The projects are coded as C026-54(5) and F-54-3(14) by the Missouri State Highway Commission. Each set of records encompassed the full duration of the construction of a section of two-lane, hard surface highway.

12 in. of soil and is based upon technical data supplied by the U.S. Forest Service and the U.S. Army Corps of Engineers. For purposes of a cursory investigation of the empirical results, it is sufficient to indicate that the index is always positive and takes high values when the soil is wet and low values when the soil is dry. Maunder et al. (1971) describe the derivation of the soil moisture index. As for the other variables, temperatures are measured in degrees Fahrenheit and precipitation is measured in hundredths of an inch.

4. ESTIMATED RELATIONS OF CLIMATIC VARIABLES TO WORKING CONDITIONS FOR COMMON EXCAVATION

As was indicated in the preceding section, the largest number of sample observations existed for common excavation. In this section, the values of the qualitative variables reflecting work conditions for common excavation are related to the climatic variables using a number of functional forms. Since this estimation is of an exploratory nature, coefficients for each of the functional forms are reported. In addition to providing alternatives for forecasting working days, it is hoped that the forms may give some guidelines for further applications of such probability models; that is, the alternative functional forms together may give some indication as to robustness of the underlying technical relationship.

The equations were estimated using the generalized least squares, normit, and logit procedures. The results are included in tables 1, 2, and 3, respectively. In each case, the information presented consists of estimates of the parameter values on the included variables and a measure of the error.[7] The measure of the error can be used to evaluate the comparative forecasting accuracy of the equations listed. As the discussion in section 2 would imply, the equations for each of the estimation methods are linear in the parameters.

Because of the forecasting limitations of the generalized least-squares estimations and the computational difficulties associated with the normit estimations, tables 1 and 2 include fewer functional forms than table 3. We only include enough equation estimates to allow some comparison with the logit results in table 3. With the limitations discussed in section 2, the equations recorded in tables 2 and 3 look reasonable; that is, negative coefficients on precipitation variables and small positive coefficients on temperature variables. For the generalized least-squares estimation method, eq (E) in table 1 appears to give the best fit and, thus, the best predictive ability. Equation (A), which has only the soil moisture index and average temperature as arguments, also has some appeal on the basis of its simplicity.

For the normit results, eq (A) appears to give the best fit (table 2). Note also that eq (A) is a simple linear function in the soil moisture index. The comparative fits of the estimated functions for the normit and generalized least-squares methods suggest that simple equations involving precipitation are likely to do relatively well, as compared to more complicated polynomials in precipitation and temperature, as predictors of working day probabilities for common excavation. It is also apparent that some physically based function in precipitation, such as the soil moisture index, can be effectively employed as a forecasting variable.

The logit equations (table 3) are more varied than either the normit or generalized least-squares results. Our preliminary results suggest that the logit method produced the best fits, and because of its computational advantages, a number of alternatives not previously considered were explored. The most interesting of these, eq (G) in table 3, included 3-day average precipitation, the log of average daily temperature, and the 0, 1 variable indicating whether or not precipitation occurred on the day in which y_n was observed. The small error, correct sign pattern, and possibilities of furnishing probabilities for the 0, 1 variable for purposes of forecasting seem to designate it as the most promising of any of the equations estimated by the three methods.[8]

In summary, we conclude that the probability models can be effectively applied to the physical relationships involving climate variables and work day probabilities. The coefficients estimated on precipitation variables were consistently negative and those on temperatures were positive. Generally, the success of the application suggests that some of the work-no work decisions can possibly be made in a quantitative and impersonal manner. We illustrate the best of these models in the following section.

5. AN ILLUSTRATIVE APPLICATION OF THE PROBABILITY MODELS

The equations presented in tables 1–3 may be difficult to evaluate with respect to their actual predictive ability. For this reason, we have elected to illustrate the probability models by comparing the predicted values to those that actually occurred in the sample period. Two months are selected for this purpose; one more wet than typical and one more dry than typical. The extreme periods were selected so that the predictions from the probability models could be compared under the least favorable conditions. (These are sample extremes.) In addition to the observed values of the dichotomous dependent variable and the concomitant values of the included climatic variables, a number of special comments are included. These comments are recorded on the days in which the predictions from the probability model do not appear to conform with the observed value of y_n. These comments may give some insight into possibilities for more refined specifications of functional forms for such models as well as indicating the limitations of those presented.

[7] The measure is just the error sum of squares. A variable for the regressions could be calculated simply by dividing the number by 218, less the number of parameters estimated in each equation.

[8] Further exploration of the applicability of this equation would be a very useful exercise, particularly as it relates to the decision aspects of the model. Weather information involving precipitation is frequently given in terms of probabilities. Hence, the work-no work probability is jointly determined by the error of the statistical relationship describing the probability model and the precipitation probability. Although calculation of exact results for such situations may be cumbersome, a numerical application of this result using simulation modeling would be feasible.

499-057 O - 73 - 6

Table 1.—*Generalized least-squares estimates of linear equations relating climatically oriented variables to working day probabilities for common excavation. Numbers in columns refer to the regression coefficients of the variables heading the respective columns.*

Equation	X_{0n}	X_{1n}	X_{2n}	X_{3n}	X_{2n}^2	X_{6n}^2	X_{6n}	$\ln X_{1n}$	$X_{2n}X_{6n}$	X_{3n}^2	Error
A	1. 1947	−0. 4296									29. 17
B	0. 7341							−0. 3492			32. 38
C	1. 1126		−6. 1571		6. 2250	0. 00004	−0. 0043		0. 0323		28. 05
D	0. 7707			−1. 3634		0. 00003	−0. 0002				27. 40
E	0. 8308			−3. 5582			0. 0020			3. 7844	24. 63

$X_{0n}=1$
X_{1n}=soil moisture index
X_{2n}=7-day average precipitation (10^{-2} in.)
X_{3n}=4-day average precipitation (10^{-2} in.)
X_{6n}=average temperature (°F)

Table 2.—*Normit estimates of linear functions relating climatically oriented variables to working day probabilities for common excavation. Numbers in columns refer to the regression coefficients of the variables heading the respective columns.*

Equation	X_{0n}	X_{1n}	$\ln X_{1n}$	X_{2n}	X_{6n}	$\ln X_{2n}$	$\ln X_{6n}$	$X_{2n}X_{6n}$	X_{6n}^2	X_{2n}^2	Error
A	4. 5685	−3. 3123									25. 98
B	0. 9230		−3. 0718								27. 89
C	−0. 6869			−0. 0860	0. 0339						31. 80
D	7. 4842					−0. 8408	2. 2537				30. 30
E	3. 4260			−0. 2580	−0. 0663			0. 0006	0. 0007	0. 0034	30. 90

$X_{0n}=1$
X_{1n}=soil moisture index
X_{2n}=7-day average precipitation (10^{-2} in.)
X_{6n}=average temperature

Table 3.—*Logit estimates of linear functions relating climatically oriented variables to working day probabilities for common excavation. Numbers in columns refer to the regression coefficients of the variables heading the respective columns.*

Equation	X_{0n}	X_{1n}	X_{2n}	X_{6n}	X_{2n}^2	X_{6n}^2	$X_{2n}X_{6n}$	X_{3n}	X_{3n}^2	$\ln X_{6n}$	X_{4n}	X_{4n}^2	X_{5n}	X_{7n}	Error
A	8. 352	−5. 5980													22. 28
B	−0. 104		−0. 1031	0. 0374											31. 12
C	3. 798		−0. 3285	0. 0484	0. 0036	0. 0005	0. 0015								28. 65
D	13. 545							0. 5232	0. 0095	3. 7840					22. 83
E	−5. 128		−0. 0242		0. 0037					1. 9065					28. 69
F	−13. 366									3. 8237	0. 0640	0. 0002			19. 67
G	−7. 029									2. 3464			3. 0658	0. 0613	23. 78

$X_{0n}=1$
X_{1n}=soil moisture index
X_{2n}=7-day average precipitation (10^{-2} in.)
X_{6n}=average temperature
X_{3n}=4-day average precipitation (10^{-2} in.)
$X_{4n}=X_{7n}+$ current day's precipitation (10^{-2} in.)
X_{5n}=0, 1 variable daily precipitation
X_{7n}=3-day average precipitation (10^{-2} in.)

The wet month within the sample period was April 1966. Table 4 includes the sample observations for this month together with the comments and the estimated work probabilities from models F and G in table 3. A comparison of the predicted values with the observed variable y_n indicates that the models give the largest error on days in which late afternoon rains occur. Although model G does better than model F, it is obvious that an alternative specification could produce more precision in the forecasts. This lack of precision makes it difficult to decide whether or not to charge for a work day. One could, however, establish critical levels of work day probabilities to give fairly accurate decisions as to the charging and planning of work days.

The dry month selected for discussion was July 1966. Predictions from logit models F and G are also reproduced for this dry month in table 5.

The only no-work day in the month occurred on July 26. Although an ideal equation might estimate the probability more closely, the two models we have selected do rather well. Comparison of the dry and wet months suggests that, given the two estimated functional forms, the threshold or critical values could be advantageously set higher for midsummer months than for early spring and late

TABLE 4.—*Sample observations and predicted work probabilities for common excavation during April 1966*

Date	Work/no-work y_n	Estimated probabilities Model F	Estimated probabilities Model G	Avg. daily temperature X_{6n}	3-day avg. precipitation X_{7n}	0, 1 precipitation variable X_{5n}	Special comments
1	1	0.938	0.945	67.0	0	0	
4	1	.687	.840	40.5	0	0	
5	1	.687	.840	40.5	0	0	
6	1	.707	.847	41.5	0	0	
7	1	.767	.870	45.5	0	0	
14	0	.181	.263	46.0	0.497	1	
15	1	.425	.572	48.0	.293	0	
18	0	.574	.379	63.0	.360	1	Rain in late afternoon
19	0	.569	.273	68.0	.443	1	Rain in late afternoon
20	0	.018	.122	57.0	2.377	1	
22	0	.070	.553	50.5	0.833	0	
25	1	.823	.859	61.5	.137	0	
26	0	.810	.225	62.0	.157	1	Rain beginning after 1400 LST
27	0	.310	.393	62.5	.567	1	
28	1	.596	.141	57.5	.293	1	Rain, previous evening
29	1	.714	.767	60.5	.233	0	

TABLE 5.—*Sample observations and predicted work probabilities for common excavation during July 1966*

Date	Work/no-work y_n	Estimated probabilities Model F	Estimated probabilities Model G	Avg. daily temperature X_{6n}	3-day avg. precipitation X_{7n}	0, 1 precipitation variable X_{5n}	Special comments
1	1	0.966	0.962	79.0	0	0	
5	1	.973	.967	84.0	0	0	
6	1	.975	.968	86.0	0	0	
7	1	.974	.968	85.0	0	0	
8	1	.965	.961	78.5	0	0	
11	1	.975	.968	85.5	0	0	
12	1	.980	.972	91.0	0	0	
13	1	.978	.970	88.5	0	0	
14	1	.976	.969	86.5	0	0	
15	1	.976	.969	86.5	0	0	
18	1	.955	.942	88.0	0.1133	0	
19	1	.955	.939	86.0	.0131	0	
20	1	.951	.964	81.0	0	0	
21	1	.958	.956	74.5	0	0	
22	1	.952	.953	72.0	0	0	
25	1	.866	.543	79.5	0	1	0400 LST rainfall
26	0	.883	.419	79.5	0.0833	1	Rainfall about noon
27	1	.933	.924	81.0	.1300	0	
28	1	.939	.928	83.0	.1300	0	
29	1	.694	.431	84.0	.0467	1	Heavy rainfall late afternoon
31	1	.632	.603	73.5	.4333	0	

fall months. In any large-scale application of this procedure, however, one would probably break the sample period into subperiods as a means of handling the difficulties suggested by tables 4 and 5.

6. APPLICATIONS OF PROBABILITY MODELS TO OTHER CONSTRUCTION ACTIVITIES

In section 3, we noted that the records from which the information on common excavation were obtained contained data on other construction activities. In this section, data from two of these activities are employed to estimate probability functions. Our purposes in presenting these results are to illustrate the applicability of the probability model to other types of construction activities and to provide a basis for evaluating the comparative effects of climatic variables on other activities.

The two activities selected for the application are (1) paving and asphalt, and (2) bridge, culvert, and drainage structures. The regression equations that were estimated are of the form

$$w_n=-15.5157+0.0037X_{4n}-0.0004X_{4n}^2+4.1588\ \ln X_{6n} \tag{22}$$

for asphalt and paving and

$$w_n=14.8786-69.8991X_{4n}+31.5216X_{4n}^2+3.9810\ \ln X_{6n} \tag{23}$$

for bridge, culvert, and drainage structures where

$$\text{Prob.}\,(y_n=1|X_n')=\frac{e^w}{1+e^w}.$$

Both models were estimated by the logit method, and the errors for the models are comparable to those for common excavation.

These results appear to be acceptable. For asphalt and paving, the variable X_{4n}, which reflects the precipitation for 4 days, is far less important, in terms of decreasing work day probabilities, than for either the common excavation equations or the bridge, culvert, and drainage structures equation. In the latter case, the precipitation variable is important, as would be expected. Average temperatures, X_{6n}, are of approximately equal importance in both equations. This may at first appear to be questionable, but when it is recalled that the data are only for the summer season, the result is reasonable.

7. SUMMARY AND CONCLUSION

We have suggested and applied probadility models to the problem of determining working days for road construction activities. The results of the application are encouraging. A number of functions that are simple in terms of their arguments involving climatic variables appear to fit the data rather well. In fact, the accuracy achieved with this limited amount of data indicates that a more thorough application of these methods could provide a substantive basis for planning by highway departments and construction firms. These plans could be a basis for contract penalty clauses and other technical specifications as well as a basis for long-range projections relating to equipment requirements and payment schedules. In many instances, these types of decisions are already being made with the aid of data provided by simulation models. These probability functions could be easily incorporated into such planning models as a method of portraying climate-oriented uncertainties.

ACKNOWLEDGMENT

Research for this paper was supported by National Science Foundation Grant GA15413. The authors also wish to acknowledge Carl Klamm, Construction Division engineer for the Missouri State Highway Commission, and the Missouri State Highway Commission for their cooperation in providing the data used in the paper and for helpful comments on various aspects of the problem.

REFERENCES

Berkson, J., "Estimate of the Integrated Normal Curve by Minimum Normit Chi-Square With Particular Reference to Bio-Assay," *Journal of the American Statistical Association*, Vol. 50, No. 269, Washington, D.C., June **1955**, pp. 529–549.

Cox, David Roxbee, "Some Procedures Connected With the Logistic Qualitative Response Curve," *Research Papers in Statistics*, John Wiley & Sons, Inc., London, England, **1966**, pp. (See pp. 55–71.)

Goldberger, Arthur Stanley, *Econometric Theory*, John Wiley & Sons, Inc., New York, N.Y., **1964**, 399 pp.

Maunder, W. John, Johnson, Stanley R., and McQuigg, J. D., "Study of the Effect of Weather on Road Construction: A Simulation Model," *Monthly Weather Review*, Vol. 99, No. 12, Dec. **1971**, pp. 939–945.

Tobin, J., "The Application of Multivariate Probit Analysis to Economic Survey Data," *Cowles Foundation Discussion Paper* 1, Cowles Foundation, New Haven, Conn., **1955**, 48 pp.

Zellner, A., and Lee, T. H., "Joint Estimation of Relationships Involving Discrete Random Variables," *Econometrica*, Vol. 33, No. 2, Econometric Society, New Haven, Conn., Apr. **1965**, pp. 382–394.

[Received April 20, 1972; revised October 17, 1972]